

**A**

**MINI-PROJECT REPORT**

**ON**

**“Easy Power Predictor”**

**Computer Science and Engineering**

**Walchand Institute of Technology**

**(An Autonomous Institute)**

By

**Shreya Gandhi TECSE A-45**

**Vipul Ghodke TECSE A-46**

**Nisha Gurav TECSE A-47**

**Shreyansh Kahate TECSE A-48**

Under Guidance Of

**Mrs. K.A. Khedikar**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**WALCHAND INSTITUTE OF TECHNOLOGY**

**SOLAPUR - 413006**

**(2023-2024)**



**CERTIFICATE**

This is to certify that the Mini-Project entitled

**“Easy Power Predictor”**

Is

Submitted By

**Shreya Gandhi TECSE A-45**

**Vipul Ghodke TECSE A-46**

**Nisha Gurav TECSE A-47**

**Shreyansh Kahate TECSE A-48**

**(Mrs. K.A. Khedikar) (Dr.Mrs.A.M.Pujar)**

***Project Guide Head***

**Dept of Computer Sc. & Engg**

**(Dr. V .A.Athavale)**

**Principal**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**WALCHAND INSTITUTE OF TECHNOLOGY**

**SOLAPUR**

**(2023-2024)**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr No.** | **Topic** | **Page No.** |
| 1. | Abstract | 4 |
| 2. | Introduction | 5 |
| 3. | Problem Statement | 6 |
| 4. | Objectives | 6 |
| 5. | Background | 7 |
| 6. | Technologies Used | 9 |
| 7. | System Architecture & Working | 10 |
| 8. | Implementation | 11 |
| 9. | Testing Report | 16 |
| 10. | Results and Screenshots | 17 |
| 11. | Advantages and Disadvantages or Applications | 20 |
| 12. | Future scope | 21 |
| 13. | Overall Project Cost | 22 |
| 14. | Conclusion/Summary | 23 |
| 15. | References | 25 |

**Abstract**

"Easy Power Predictor" is a machine learning model designed to forecast electricity consumption by integrating various weather and holiday factors. The objective is to enhance electricity distribution efficiency, reducing losses, and increasing cost-effectiveness by supplying optimal power levels. The dataset, sourced from Kaggle, underwent thorough preprocessing, merging, and cleaning to integrate data from diverse sources, including weather and holiday data. The project utilized machine learning algorithms such as Linear Regression, Random Forest, and XGBoost for prediction tasks. These models were trained, validated, and evaluated using different data subsets, employing metrics like Mean Absolute Error , R-squared Value , Mean Squared Error and Root Mean Squared Error to assess performance. Through a combination of visual analysis and rigorous model evaluation, the project identified Random Forest as the most effective model for accurate electricity consumption forecasting. This predictive tool holds significant potential for optimizing power distribution strategies, reducing energy costs, and enhancing overall efficiency in electricity management.

**Introduction**

The "Easy Power Predictor" project aims to address the complexities of electricity production and consumption dynamics in today's energy landscape, particularly with the rising integration of renewable energy sources. By leveraging machine learning techniques, this project seeks to develop a predictive model that accurately forecasts energy demand, bridging the gap between electricity supply and demand. The predictive model considers various influencing factors such as weather conditions, time of day, and holiday indicators, which play pivotal roles in determining electricity usage.

To achieve this, the project begins with comprehensive data acquisition and processing from diverse sources like smart meter readings, weather data, and holiday schedules. The initial raw data, often unstructured and disorganized, undergoes thorough cleaning and transformation to create a coherent dataset suitable for analysis. This involves handling missing values, transforming date and time data, and merging data from different sources to form an integrated dataset.

Following data preparation, the project employs exploratory data analysis (EDA) to uncover patterns, correlations, and trends in the data. Visualizations like histograms, scatter plots, and correlation heatmaps offer insights into the relationships between different variables, shedding light on the impact of weather and holidays on electricity consumption.

With a refined dataset ready for modeling, the project explores various machine learning algorithms, including Linear Regression, Random Forest, and XGBoost, to forecast energy usage. Each model is trained and evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared values. Through feature engineering and model development, the project aims to optimize energy management strategies, promising proactive resource allocation, reduced energy wastage, and enhanced efficiency in utility operations.

Ultimately, the project's goal is to identify the most effective model for predicting electricity consumption, providing power providers with valuable insights to make informed decisions about energy management and distribution. The outcomes of the "Easy Power Predictor" project highlight its significance in the context of energy optimization, sustainability, and the efficient distribution of electricity to various regions, ensuring reliable and cost-effective energy supply.

**Problem Statement**

The production and consumption of electricity face significant challenges due to fluctuating demand and supply, exacerbated by the growing adoption of renewable energy sources. This leads to instances of overproduction, energy wastage, and occasional shortages in meeting electricity demand. There is a need for accurate forecasting tools to proactively manage energy resources and bridge the gap between supply and demand.

**Objectives**

* To reduce energy wastage by improving the accuracy of demand forecasting, leading to more effective energy distribution and consumption.
* To identify relevant features impacting electricity consumption through comprehensive feature engineering.
* To evaluate model performance using appropriate metrics like Mean Absolute Error (MAE) , Mean Squared Error (MSE) and Root Mean Squared Error (MSE).

**Background**

The global energy landscape is rapidly evolving with an increasing shift towards sustainable and renewable energy sources. This transition brings new challenges to the efficient distribution and management of electricity due to the intermittent nature of renewables and dynamic consumption patterns influenced by factors like weather and holidays.

Traditional methods of electricity generation are being supplemented by renewable energy sources, leading to a more complex energy grid. Accurate forecasting of energy demand becomes crucial for optimizing distribution, minimizing wastage, and ensuring grid reliability.

The "Easy Power Predictor" project aims to address these challenges by leveraging machine learning to develop a predictive model for electricity consumption. This model seeks to accurately forecast demand based on various influencing factors, supporting power providers in making informed decisions about energy management and distribution, ultimately contributing to a more sustainable and efficient energy future.

* Literature Review

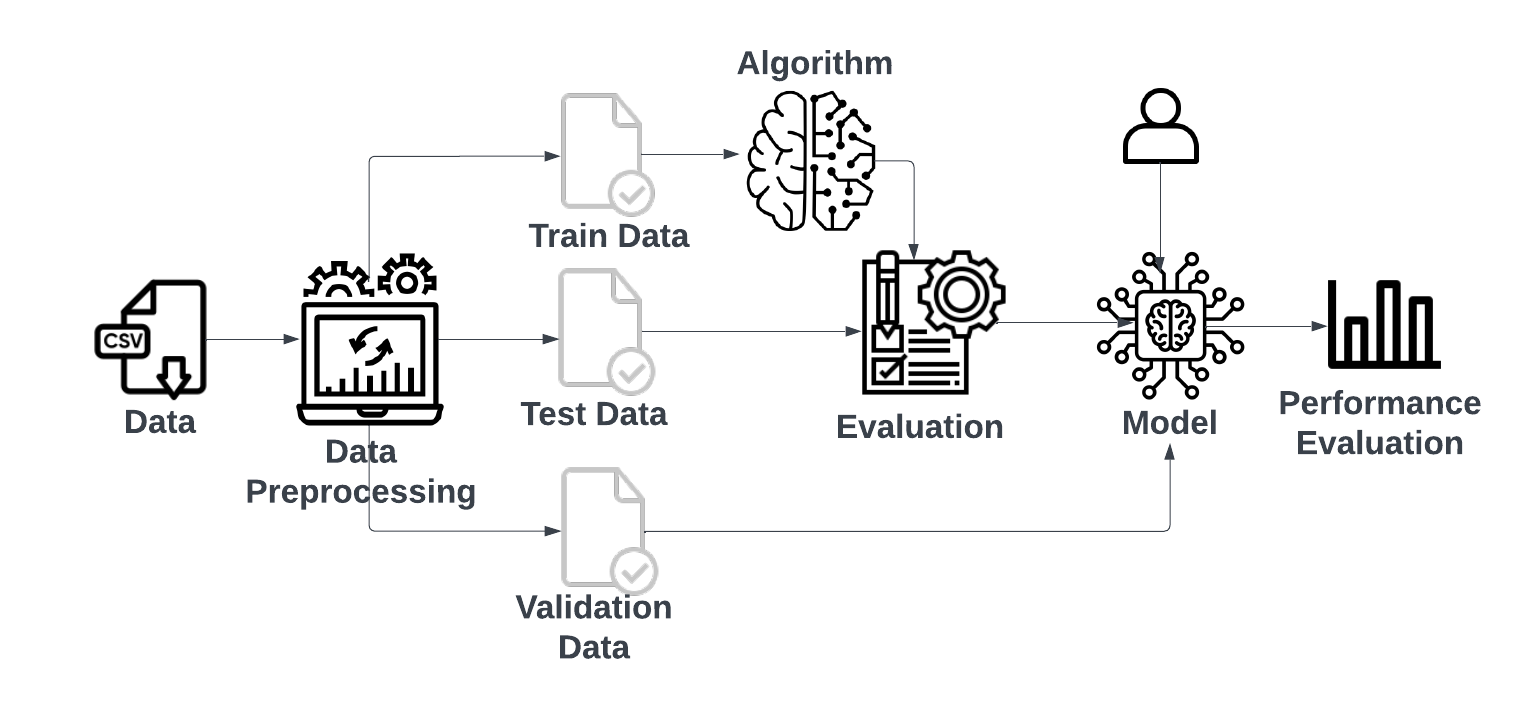
1. Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting : The paper underscores the critical role of electricity in a country's development, particularly in India, and the necessity for accurate demand forecasting. It outlines different forecasting timeframes—long-term, medium-term, and short-term—each serving specific planning needs. Challenges include the inability to store electricity efficiently once generated. Various machine learning algorithms have been employed, but neural networks, though promising, face issues like slow learning rates. To address these, a hybrid model combining Empirical Mode Decomposition (EMD) and Long Short Term Memory (LSTM) networks is proposed for better forecasting accuracy.
2. Investigation of Performance of Electric Load Power Forecasting in Multiple Time Horizons with New Architecture Realized in Multivariate Linear Regression & Feed-Forward Neural Network Techniques :The paper highlights the significance of Electric Load Power Forecasting (ELPF), particularly short-term load forecasting (STLF), for efficient power system operation. It discusses the evolution of forecasting techniques from linear regression to artificial intelligence-based methods like neural networks. The paper proposes a comparative analysis between statistical regression and neural network techniques for load forecasting, emphasizing the importance of weather variables. It outlines the hardware and software used for model development and provides an overview of the paper's organization, including sections on data processing, theoretical backgrounds, implementation, results, conclusion, and future directions.
3. Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market : The paper discusses the challenges of accurately forecasting electricity consumption in Algeria, specifically focusing on the economic sector and high-voltage consumers. It introduces the objective of developing ensemble models using deep learning techniques to address these challenges, aiding SONELGAZ's planning department in electricity sales and purchases. The structure of the paper includes a review of related work, methodology description, results presentation, and concluding insights.
4. Electric Heating and the Effects of Temperature on Household Electricity Consumption in South Africa: The paper investigates how temperature changes affect household electricity consumption in South Africa, particularly focusing on responses to cold temperatures. It finds that households use significantly more electricity in colder weather, but responses vary by season, with less impact during summer months. This suggests that daily temperature fluctuations may have less effect on energy consumption compared to overall seasonal temperature increases. The paper reviews existing literature, presents data and methodology, discusses results, and concludes with simulations of household responses to predicted temperature increases.
5. Research on Deep Learning Energy Consumption Prediction Based on Generating Confrontation Network: The paper emphasizes energy conservation and reduction, highlighting the need for accurate energy consumption prediction. It proposes using deep learning, specifically convolutional neural networks (CNNs) and generated confrontation networks (GANs), to enhance prediction models. By analyzing spatial correlations in energy data, the study aims to improve prediction accuracy. Experimental results show the effectiveness of the approach, with the combined deep learning and GAN model outperforming traditional methods.

**Technologies Used**

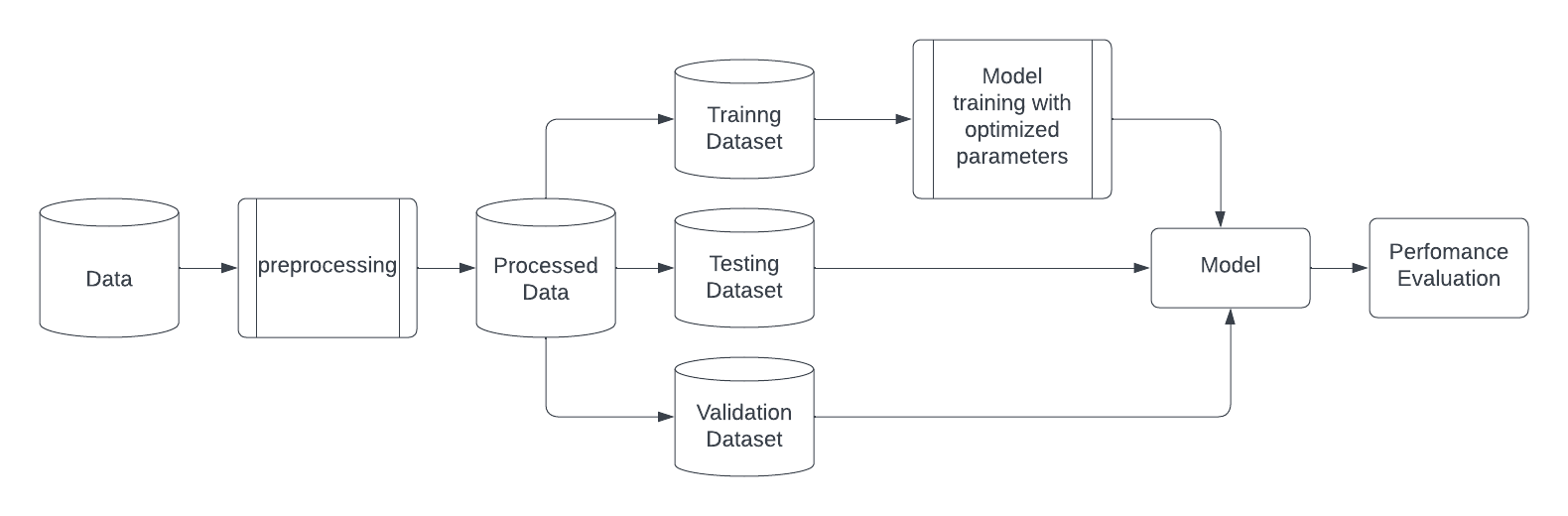
The "Easy Power Predictor" project leverages a combination of technologies and tools to develop and deploy the predictive model for electricity consumption forecasting. The key technologies used include:

* **Python:** Python serves as the primary programming language for data preprocessing, modeling, and implementation of the predictive model. Python's extensive libraries and frameworks are utilized for machine learning tasks.
* **Pandas:** Pandas is used for data manipulation and preprocessing tasks, including handling of time-series data, feature extraction, and data cleaning.
* **Scikit-Learn (sklearn):** Scikit-Learn is employed for building and training machine learning models such as linear regression, decision trees, and ensemble methods. It provides efficient implementations of various algorithms and tools for model evaluation.
* **Matplotlib and Seaborn:** These libraries are used for data visualization, allowing for the creation of informative plots and graphs to analyze relationships between variables and visualize model outputs.
* **Jupyter Notebook:** Jupyter Notebook provides an interactive environment for code development, data exploration, and model prototyping. It facilitates a seamless workflow for experimenting with different approaches and visualizing results.
* **Machine Learning Algorithms**: Various regression algorithms (e.g., linear regression, decision trees) are implemented using Scikit-Learn to build the predictive model based on historical electricity consumption data and relevant features.

**System Architecture & Working**



The architecture of the system begins with providing the data to the project. The next step involves performing all required preprocessing on the data, including cleaning and feature engineering. The data is then split into training, testing, and validation sets. Following this, the model is trained using the training data and tested to evaluate its performance. Once the initial testing is complete, the model is validated on the validation data to ensure its generalizability. The model's performance is then assessed using various evaluation metrics such as MAE, MSE, RMSE, and R-squared. Finally, a visual comparison of model performance is conducted to choose the best-performing model based on the evaluation results.



Flowchart

**Implementation**

The implementation phase of the project was characterized by iterative development and evolving methodologies. This section outlines the detailed steps and procedures followed, including data preprocessing, Exploratory Data Analysis (EDA), model development and training, evaluation, and insights gained from the project's progression.

1. Data Preprocessing

Data Cleaning:

The dataset underwent cleaning to handle missing values, and outliers were removed to ensure data quality and reliability.

Feature Engineering:

Selected relevant features such as temperature, humidity, wind speed, and others for model input.

Date-wise Energy Sum Calculation:

Aggregated daily energy consumption by summing the energy values for each date.

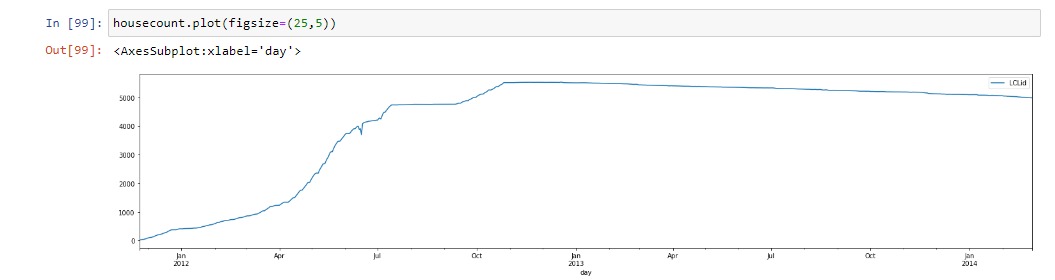
2. Exploratory Data Analysis (EDA)

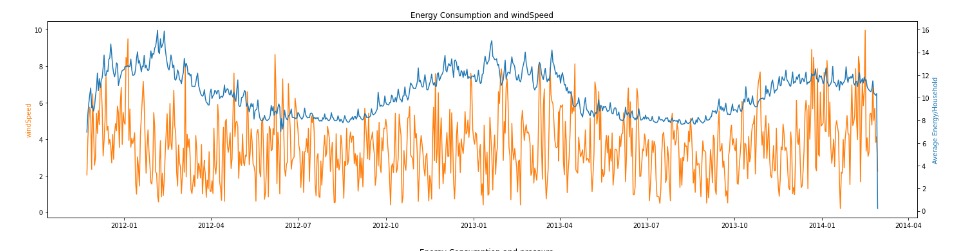
Data Visualization:

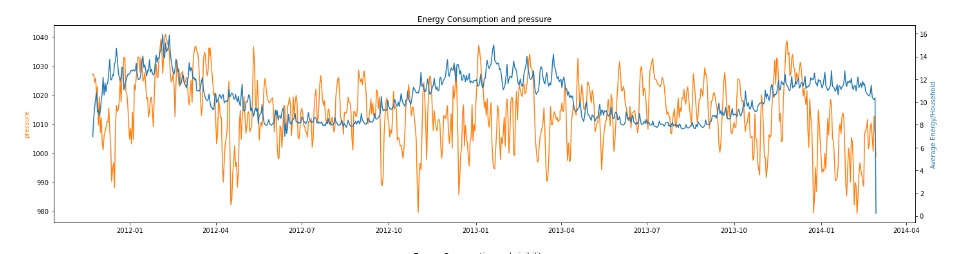
Visualized data distributions, relationships between features, and patterns in electricity consumption using graphs and charts.

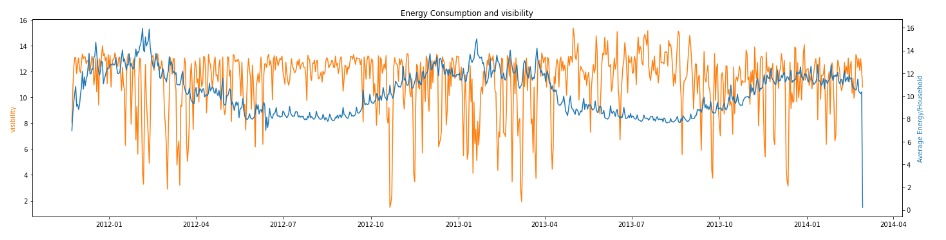
Statistical Analysis:

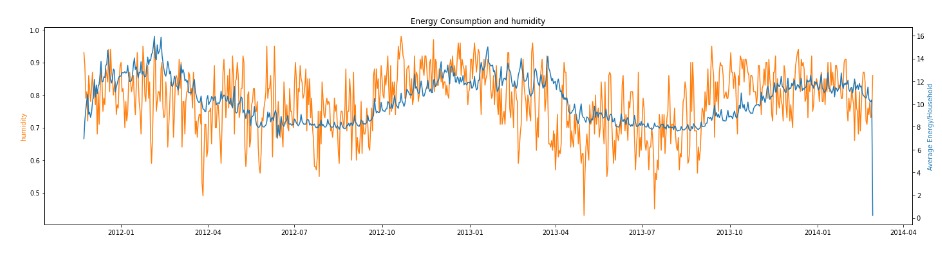
Conducted statistical tests to identify correlations between features and electricity consumption, providing insights into potential relationships and patterns.

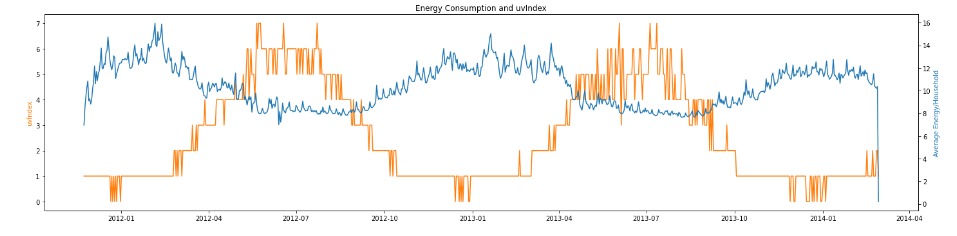


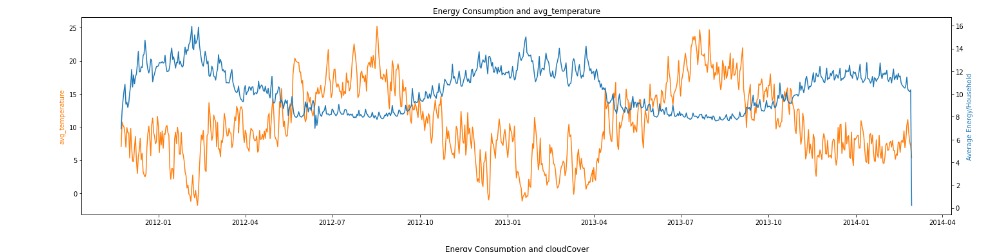


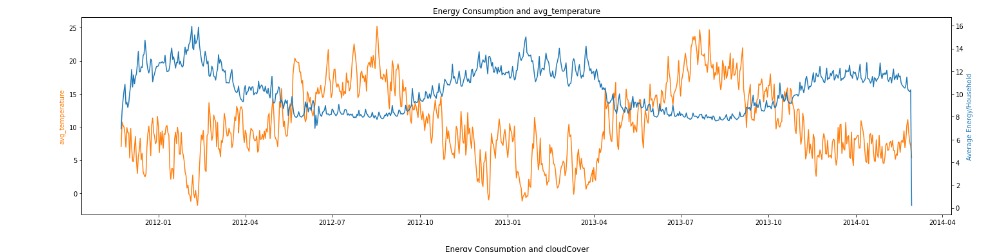


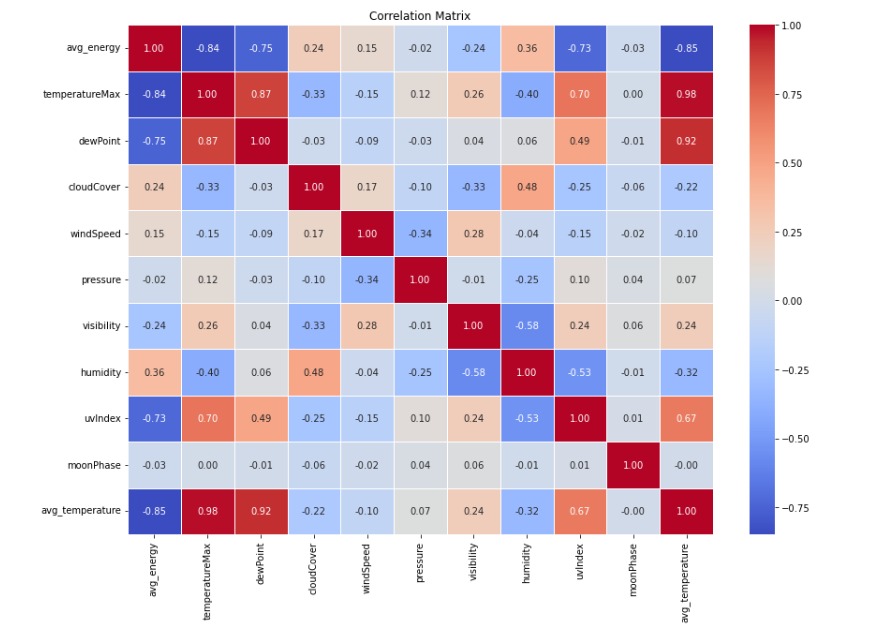


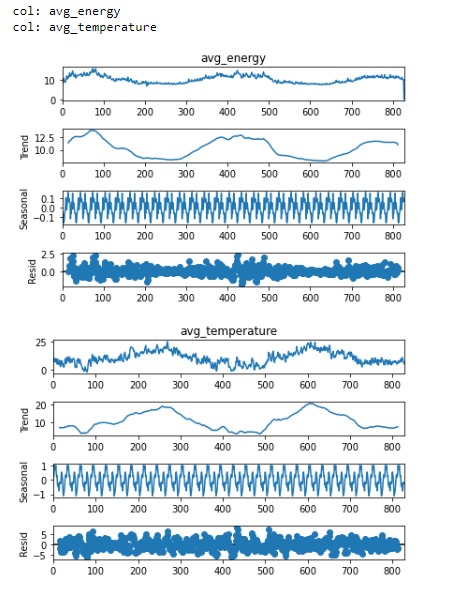












3. Model Development and Training

3.1 Individual Prediction Approach

Description:

Initially, the focus was on predicting electricity consumption for individual smart meter IDs (LCLid) using their historical data.

Challenges:

This approach faced significant challenges due to the complexity and variability in individual consumption patterns, leading to less accurate predictions.

Insights:

It became evident that the individual prediction approach was not suitable for the given dataset's characteristics and would require more sophisticated techniques like deep learning to capture the underlying patterns.

3.2 Overall Prediction Approach

Description:

To overcome the limitations of the individual prediction approach, the focus shifted to predicting the total energy consumption for a given date by considering the number of active smart meter IDs.

Advantages:

This approach aimed to capture the collective consumption pattern rather than individual variations, resulting in more stable and accurate predictions.

Model Selection:

Tested different machine learning models, including Linear Regression, Random Forest, and XGBoost, to identify the most suitable model for this approach.

4. Model Evaluation and Validation

Validation Approach:

The project employed a simple hold-out validation method, splitting the data into training, validation, and test sets, allowing effective assessment of model performance on unseen data.

Model Performance Metrics:

The models were evaluated on the test dataset using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

5. Results and Insights

5.1 Best Fit Model: Random Forest

5.2 Performance:

Random Forest has proven to be the best-performing model overall, demonstrating strong performance across all datasets (train, test, and validation).

5.3 Error Metrics:

The model maintains consistently low error metrics (Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error) and high R-squared values across all data splits.

5.4 Generalization:

Random Forest shows good generalization, avoiding overfitting while still achieving high accuracy.

5.5 Comparison with Other Models:

Although XGBoost performs exceptionally well on the training data, its performance on the test and validation data is slightly lower than that of Random Forest. This indicates a potential for overfitting in XGBoost, which may compromise its performance on unseen data.

**Testing Report**

* Dataset Split: The dataset was split into training, validation, and testing sets to ensure a comprehensive evaluation of the models' performance on different data subsets.
* Model Evaluation Metrics: The performance of the models was assessed using various metrics, including:
* Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values.
* Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values.
* Root Mean Squared Error (RMSE): The square root of MSE, providing a more interpretable measure of error.
* R-squared (R²): Indicates the proportion of variance in the dependent variable explained by the model.
* Model Performance and Findings
* Linear Regression

Training Performance: Linear Regression serves as a baseline model, providing reasonable predictions on the training data.

Testing and Validation Performance: The model performs adequately on the testing and validation data, but its simplicity may limit its ability to capture complex relationships between features and electricity consumption.

Conclusion: Linear Regression is a useful baseline model but may not be the most effective choice for this project.

* Random Forest

Training Performance: Random Forest exhibits strong performance on the training data, accurately capturing patterns in the data and avoiding overfitting.

Testing and Validation Performance: The model maintains consistent performance across testing and validation data, with low error metrics and high R-squared values.

Conclusion: Random Forest's balanced performance and generalization make it the most reliable and best-fit model for this project.

* XGBoost

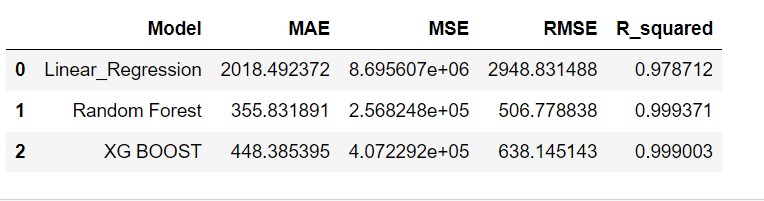
Training Performance: XGBoost demonstrates excellent performance on the training data, effectively learning complex relationships and patterns.

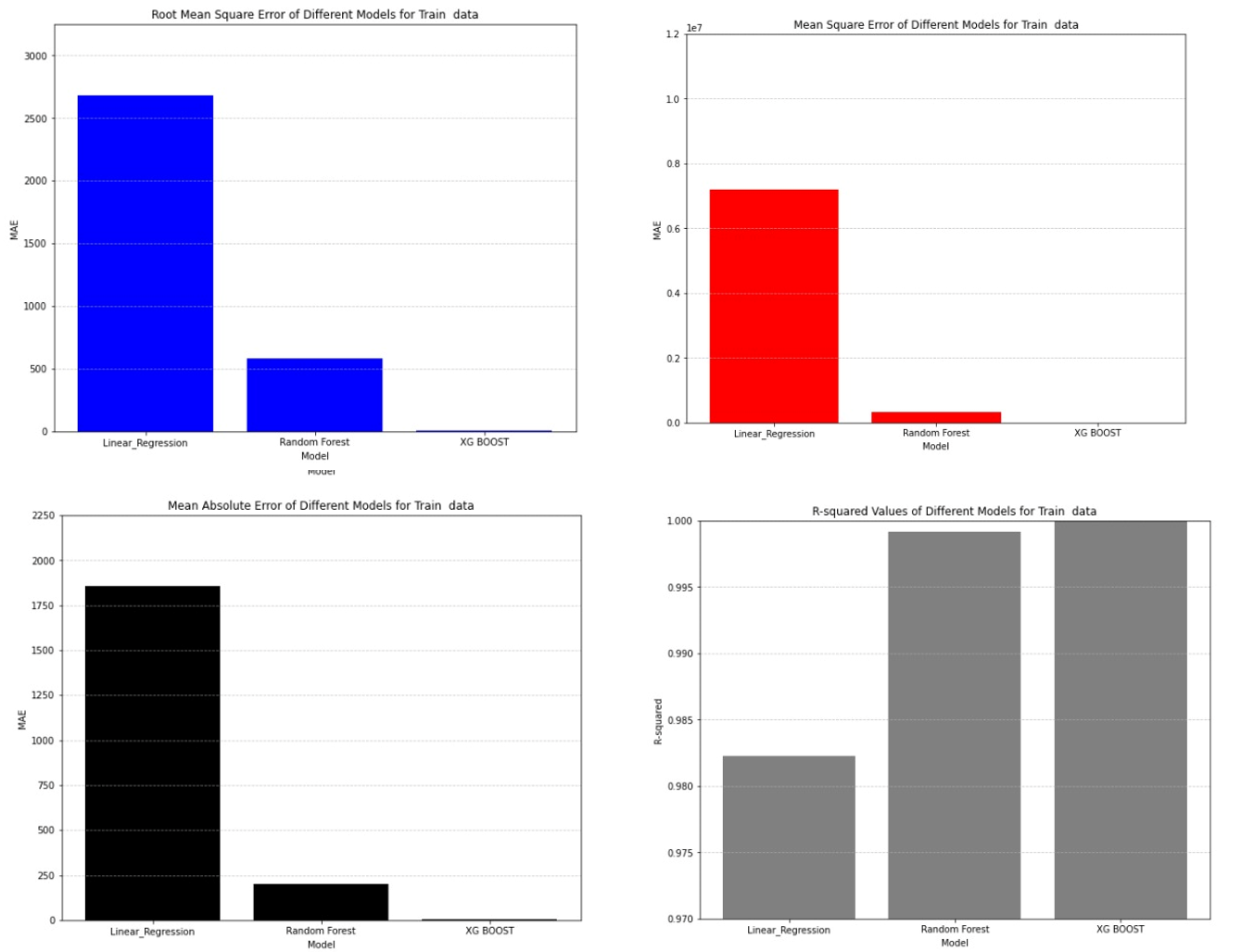
Testing and Validation Performance: While XGBoost performs well, its performance on testing and validation data is slightly lower than Random Forest, indicating a potential for overfitting.

Conclusion: Although XGBoost is a strong contender, its potential overfitting on unseen data may limit its effectiveness in this project.

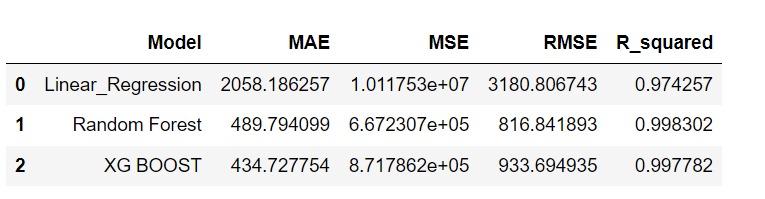
**Results and Screenshots**

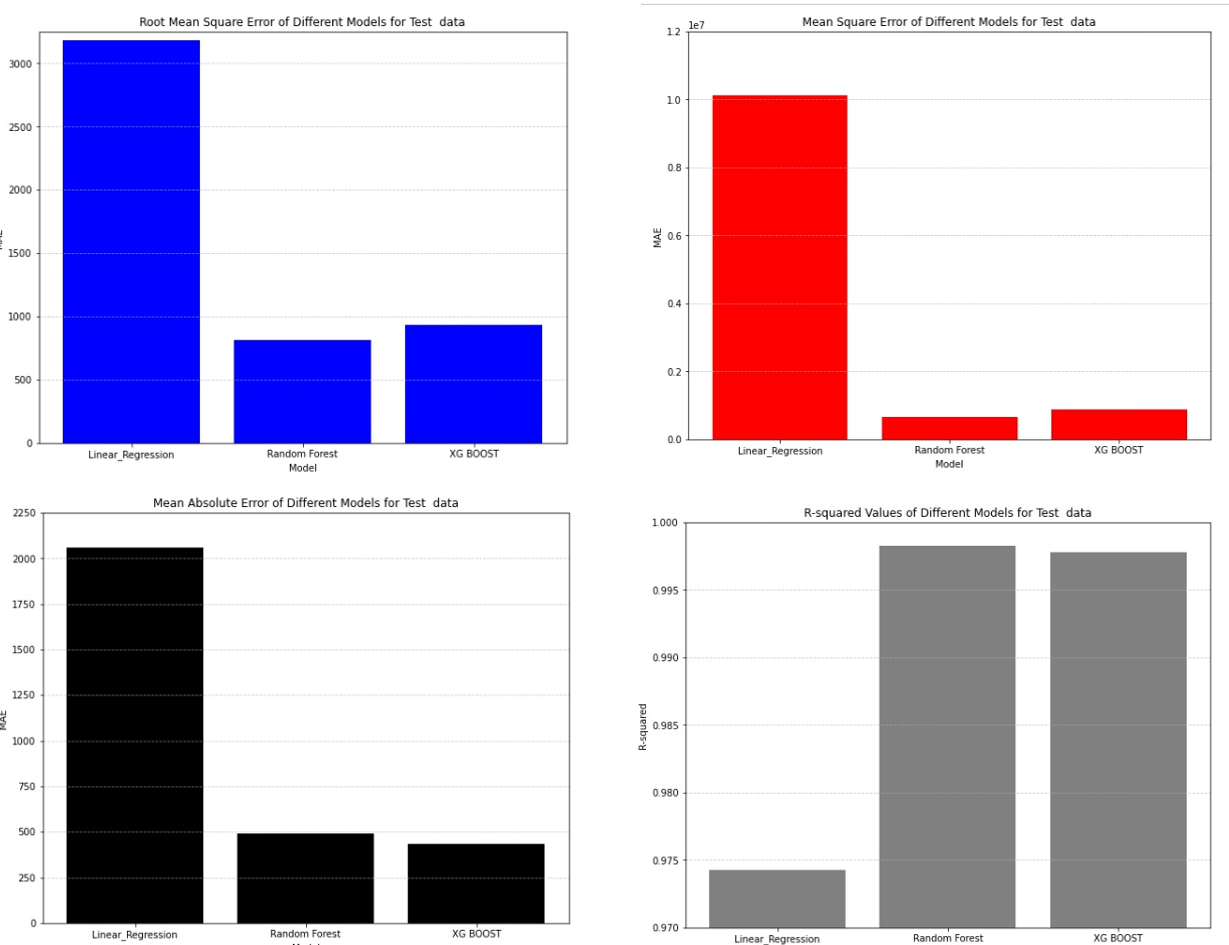
* Performance on Trained Data



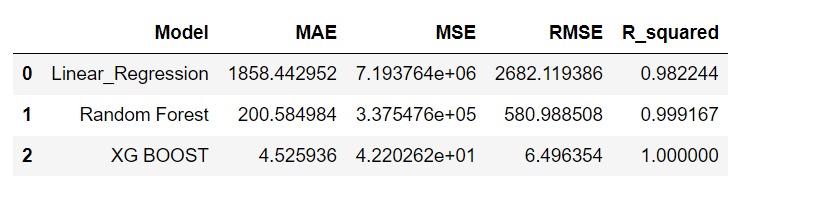


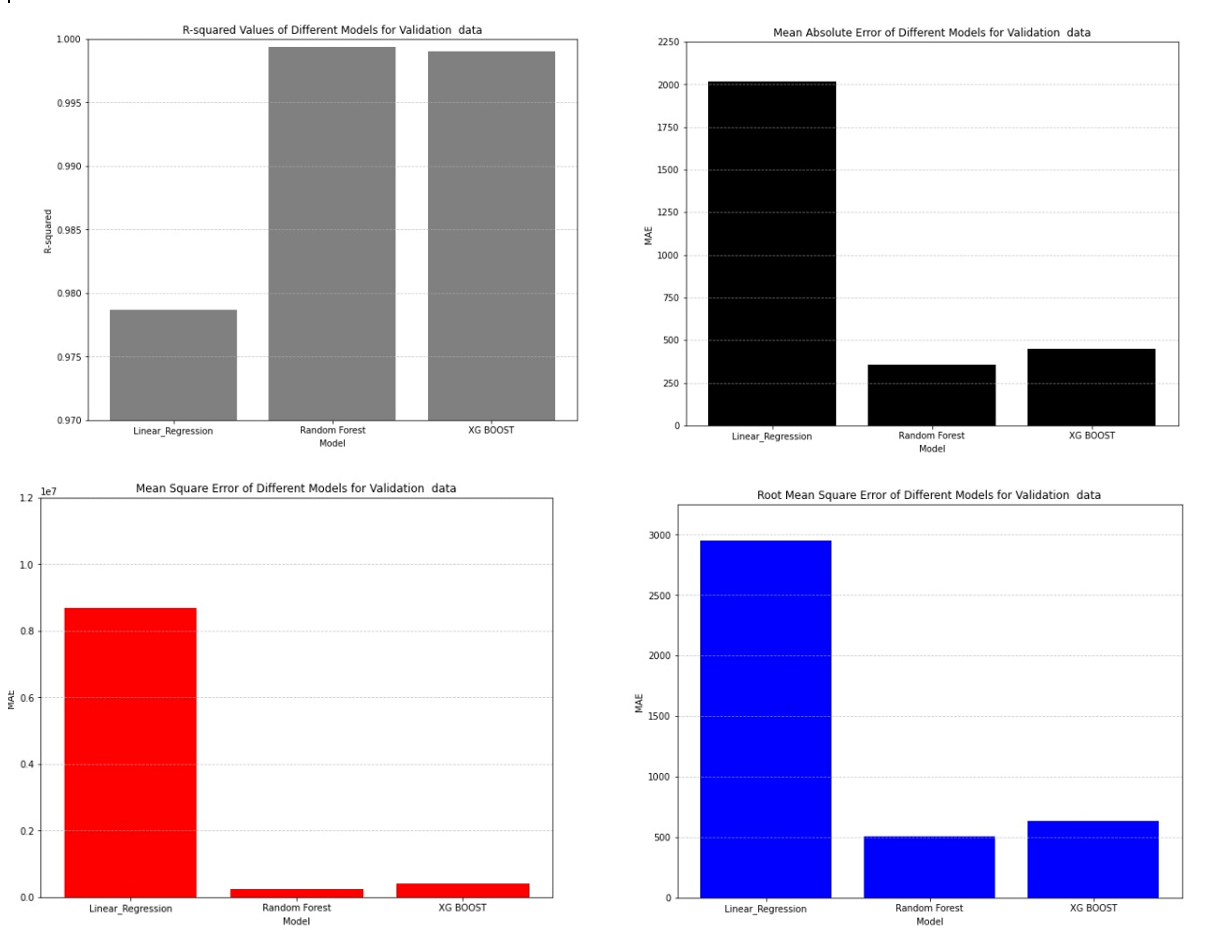
* Performance on Test Data





* Performance on Validation Data





**Advantages and Disadvantages or Applications**

Advantages:

1. Better Predictions: Machine learning considers various factors like weather and holidays, improving predictions.

2. Understanding Overall Demand: Instead of focusing on individual meters, the new approach looks at the entire network, giving a clearer picture of energy needs.

3. Adaptable to Change: The system can learn from new data and adjust, keeping predictions accurate as conditions evolve.

4. Handles Lots of Data: It can deal with large amounts of smart meter data efficiently, making it suitable for real-time use in different areas.

5. Saves Time: Automation reduces manual work, making operations smoother and more efficient.

Disadvantages:

1. Data Quality Issues: Making sure the data is accurate and consistent can be tricky, which might affect the reliability of predictions.

2. Risk of Overfitting: There's a chance the system might get too focused on specific data, making it less accurate with new information.

3. Needs Powerful Computers: Setting up and running the system might require expensive equipment, which could be a problem for some.

**Future scope**

For future work, adapting and refining the developed model for deployment in regions like India with varying meter types (e.g., household single-phase meters, industrial meters) presents an exciting challenge.

This involves:

* Data collection and preprocessing tailored to different meter types and energy usage patterns in India.
* Adjusting the model architecture and feature engineering strategies to accommodate diverse data sources and characteristics.
* Conducting rigorous testing and validation to ensure the model's effectiveness and reliability in Indian energy management contexts.
* Expanding the scope to accommodate different global contexts and metering infrastructures will enhance the applicability and impact of the developed predictive model

**Overall Project Cost**

The Easy Power Predictor incorporates a detailed evaluation of project expenses to provide a comprehensive overview of potential costs involved .This aims to offer insight into the potential financial considerations associated with executing this project.

Software:

Jupyter Notebook (Development Environment):

Jupyter Notebook is used as the primary development environment for the project, which is free and open-source.

Cost: Rs 0(no licensing cost associated)

Python (Programming Language):

Python is used for machine learning model development and implementation, and it is also free and open-source.

Cost: Rs 0 (no licensing cost associated)

Python Libraries (NumPy, pandas, scikit-learn, matplotlib):

Libraries such as NumPy, pandas, scikit-learn, and matplotlib are essential for data manipulation, modeling, and visualization.

These libraries are also free and open-source.

Cost: Rs 0 (no additional software cost)

Hourly rate:

Hourly rates are utilized to project the expenditure on manpower. Based on each team member's hourly wage, the total cost will vary according to the hours dedicated to the project throughout a 2-month period.

The project is expected to span a duration of 2 months, including planning ,development, testing phases.

Considering all the factors , Hourly rate is Rs 125,the cost of manpower can be estimated as follows:

Total Developer Hours: 5 hours/week \* 2 months = 40 hours per developer

Total Developer Costs: Total Developer Hours \* Hourly Rate \* 4 Developers

Total Developer Costs: 40\*125\*4 = Rs 20000

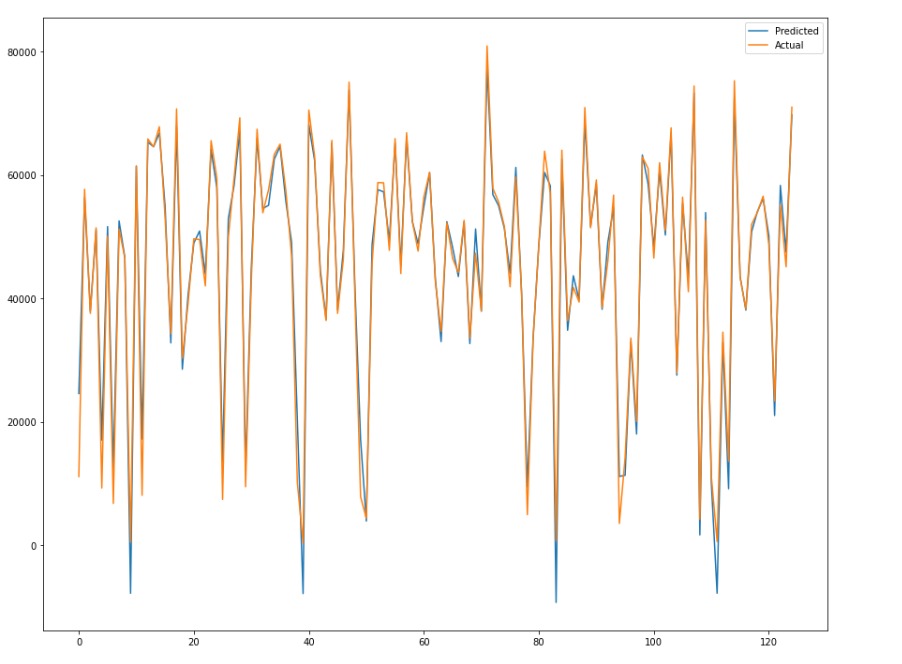
Therefore, the estimated cost of manpower for the Easy Power Predictor project is Rs 20000

**Conclusion/Summary**

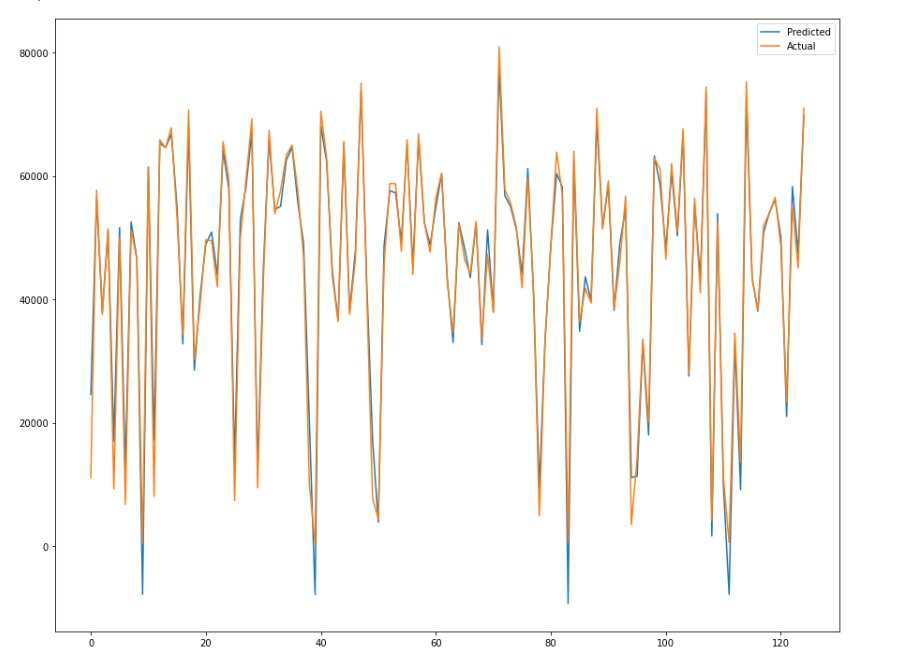
In this project, we explored predictive analysis for electricity power consumption, focusing on its relationship with weather factors and holiday data. The goal was to optimize electricity distribution and make it more cost-effective by predicting energy demand accurately. We cleaned and preprocessed data from multiple sources, including Kaggle, weather, and holiday data, to create a cohesive dataset for analysis. Through visual analysis and feature engineering, we identified key variables such as temperature, humidity, wind speed, and holiday information. We trained and evaluated various machine learning models, including Linear Regression, Random Forest, and XGBoost. Among these, Random Forest proved to be the best-performing model, demonstrating strong generalization and low error metrics across train, test, and validation datasets. This project highlights the potential of using machine learning techniques to optimize electricity distribution, minimize wastage, and achieve cost savings for utility providers.

Random Forest Fitting

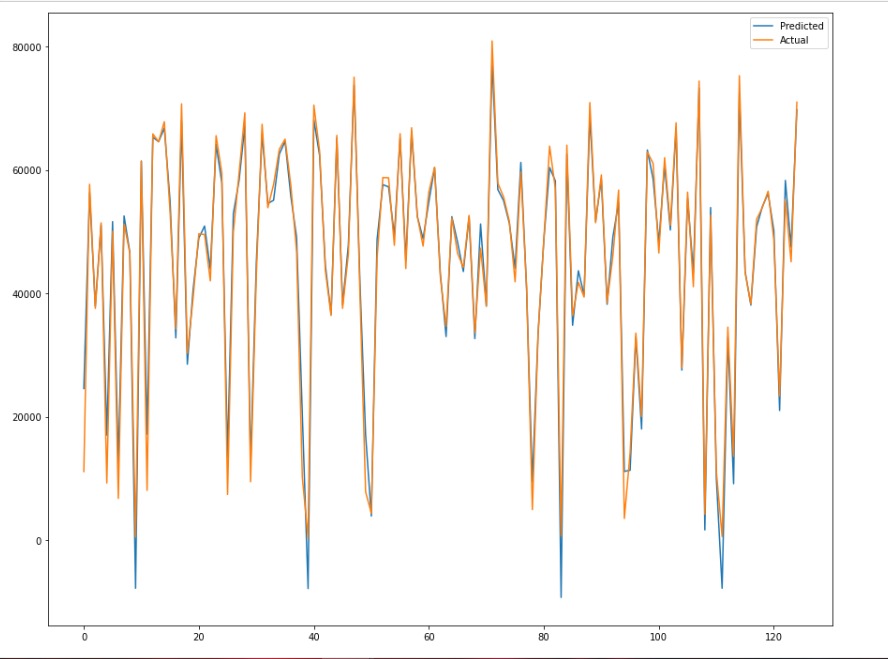
Trained data



Test Data



Validation data



**References**

1. JATIN BEDI AND DURGA TOSHNIWAL, (Member, IEEE) Department of CSE, IIT Roorkee, Roorkee 247667, India,“ Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting”,Published:August 29, 2018. Source: <https://ieeexplore.ieee.org/document/8449937>
2. M Vetri Selvi Research Scholar Electrical Engineering Department Indian Institute of Technology Delhi, India. Sukumar Mishra Professor & IEEE Member Electrical Engineering Department Indian Institute of Technology Delhi, India,“ Investigation of Performance of Electric Load Power Forecasting in Multiple Time Horizons with New Architecture Realized in Multivariate Linear Regression & Feed-Forward Neural Network Techniques”,Published in 2020

Source: <https://ieeexplore.ieee.org/document/9140425>

1. D. Hadjout a, 1 , J.F. Torres a, 1 , A. Troncoso a , A. Sebaa b, c , F. Martínez-Alvarez a, \* a Data Science and Big Data Lab, Universidad Pablo de Olavide, ES-41013, Seville, Spain b LIMED Laboratory, Faculty of Exact Sciences, University of Bejaia, 06000, Bejaia, Algeria c Higher School of Sciences and Technologies of Computing and Digital, Amizour, 06000, Bejaia, Algeria,“ Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market”,Published in 2021

Source: :<https://www.sciencedirect.com/science/article/pii/S0360544221033090>

1. SHILONG FAN School of Educational Science, Jilin Normal University, Changchun 10203, China,“ Research on Deep Learning Energy Consumption Prediction Based on Generating Confrontation Network”,Published in October 23,2019

Source: <https://ieeexplore.ieee.org/document/8880551>

1. Susanna Berkouwer,“ Electric Heating and the Effects of Temperature on Household Electricity Consumption in South Africa”,Published on February 5, 2019\

Source: <https://haas.berkeley.edu/wp-content/uploads/WP299.pdf>